

pKa measurements for the SAMPL6 prediction challenge for a set of kinase inhibitor-like fragments

Mehtap Işık^{1,2}, Dorothy Levorse³, Ariën S. Rustenburg^{1,4}, Heather Wang⁵, David Mobley⁶, Timothy Rhodes³, John D. Chodera^{1*}

¹Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States; ²Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065, United States; ³Merck & Co., Inc., MRL, Pharmaceutical Sciences, 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States; ⁴Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065, United States; ⁵Merck Co., Inc., MRL, Process Research & Development, 126 East Lincoln Avenue, Rahway, New Jersey 07065, United States; ⁶Department of Pharmaceutical Sciences and Department of Chemistry, University of California, Irvine, Irvine, California 92697, United States

***For correspondence:**

john.chodera@choderalab.org (JDC)

Abstract Determining the protonation state of a small molecule is a preliminary requirement for predicting its physicochemical and pharmaceutical properties, as well as interactions with protein targets using computational models. To determine the ionic state of a molecule in an aqueous solution at a certain pH it is necessary to know its acid dissociation constants (pK_as). As a part of SAMPL6 community challenge, we organized a blind pK_a prediction component to assess the accuracy of contemporary pK_a prediction methods. While inaccuracy in prediction of small molecule pK_as have potential detrimental impact on predictive physical models, predicting pK_as of drug-like molecules can be difficult due to challenging properties such as multiple titratable sites, heterocycles, and tautomerization. We limited the focus of this challenge on a subset of chemical space of drug-like molecules: 24 small molecules were selected to represent fragments of kinase inhibitors. We measured macroscopic pK_a values of SAMPL6 compounds with UV-absorbance based method with Sirius T3 instrument to construct an experimental reference dataset for the evaluation of computational pK_a predictions.

Introduction

Brief summary, origin and goal of SAMPL6 and outlook for future challenges.

Selection criteria of SAMPL6 compounds

Overview of the concept of microscopic and macroscopic pK_as.

Overview of pK_a measurement methods.

Brief outline of what this paper includes.

Methods

Compound selection and procurement

Small molecules were purchased in powder form.

FIGURE for compound filtering steps. Flow chart or funnel for illustration of compound selection and experimental measurement workflow.

SI TABLE vendor, lot numbers, SMILES. Selected compounds and their SAMPL challenge identities, eMolecules ID, supplier reported purity, LC/MS purity

FIGURE compound structures with pKa values + SEM

UV-metric pKa measurements

Experimental pKa measurements were collected using spectrophotometric pKa method with a Sirius T3 automated titrator instrument (Pion) at room temperature (25°C) and constant ionic strength. The UV-metric pKa measurement protocol of the Sirius T3 measures the change in multiwavelength absorbance in the 250-450 nm UV region of the absorbance spectrum while the pH is titrated between pH 1.8 and 12.2 to evaluate pKas [1, 2].

10 mg/ml DMSO solutions of each compound were prepared by weighing 1 mg of powder chemical with Sartorius Analytical Balance (Model ME235P) and dissolving it in 100 µL DMSO. DMSO stock solutions were capped immediately to limit hygroscopicity of DMSO and sonicated for 5-10 minutes in water bath sonicator at room temperature to ensure proper dissolution. These DMSO stock solutions were stored in room temperature up to 2 weeks. 10 mg/ml DMSO solutions were used as stock solutions for the preparation of 3 replicate samples for independent titrations: 15 µL of 10 mg/ml DMSO stock solution delivered to 4 mL glass sample vials of Sirius T3 with an electronic micropipette (Rainin EDP3 LTS 1-10 µL). The volume of delivered DMSO stock solution, which determines the sample concentration, is optimized individually for each compound to achieve sufficient but not saturated absorbance signal (targeting 0.5-1.0 AU) in the linear response region. Another limiting factor for sample concentration was ensuring that the compounds stays soluble in the whole range of pH titration range. 25 µL of mid-range buffer (14.7 mM K₂HPO₄ 0.15 M KCl in H₂O) was added to each sample transferred with a micropipette (Rainin EDP3 LTS 10-100 µL) to provide enough buffering capacity in middle pH ranges so that pH could be controlled incrementally throughout the titration.

pH is temperature and ionic-strength dependent. Heating block of Sirius T3 kept analyte solution at 25 °C throughout the dilution. Ionic-strength of the samples were adjusted by dilution in 1.5 mL ionic-strength adjusted water (ISA water, 0.15 M KCl) to the vials by Sirius T3. Analyte dilution, mixing, acid/base titration, and measurement of UV-absorbance was automated by UV-metric pKa measurement protocol of Sirius T3(Pion). pH was titrated between pH 1.8 and 12.2 with addition of acid (0.5 M HCl, 0.15 M KCl) and base (0.5 M KOH, 0.15 M KCl) targeting 0.2 pH steps between datapoints. Visual inspection of sample solutions after titration and inspection of pH-dependent absorbance shift in 500-600 nm region of UV spectra was used to verify no detectable precipitation occurred during the course of measurement. Absorbance in 500-600 nm region of UV spectra is associated with scattering of longer wavelengths of light in the presence of aggregates. For each analyte, we optimized analyte concentration, direction of titration and the range of pH titration in order to achieve solubility. The direction of titration was determined so that titration would start from the pH where the compound is most soluble: low-to-high pH for bases and high-to-low pH for acids. For compounds with insufficient solubility to accurately determine a pKa directly in a UV-metric titration, a cosolvent protocol was used [See the next section: UV-metric pKa measurement with cosolvent]. A protonation state change of titratable sites near chromophores will modulate the UV-absorbance spectra of these chromophores, allowing populations of distinct UV-active species to be resolved as a function of pH. To do this, basis spectra are identified and populations extracted via analysis of the pH-dependent multi-wavelength absorbance. The number of pKas is determined based on the quality of fit between experimental and modeled microstate pH-dependent populations.

This method is capable of measuring pKas between 2 and 12 when protonatable groups are at most

4-5 heavy atoms away from chromophores such that a change in protonation state alters the absorbance spectrum of the chromophore. We have selected compounds where titratable groups are close to potential chromophores (generally aromatic ring systems), but it is possible that our experimental results couldn't detect protonation of titratable groups distal to a UV-chromophore.

FIGURE: how potentiometric and UV-metric data looks side by side. Example UV-metric titration and data analysis.

Explain selection of one titration/sample

UV-metric pKa measurement with cosolvent

Explain what and why

Explain extrapolation

FIGURE: EXTRAPOLATION

Calculation of uncertainty in pKa measurements

Experimental uncertainties were reported as standard error of the mean (SEM) of three replicate pKa measurements.

Protonation site determination with NMR

Update or remove this section based on updates from lyke.

Quality control for chemicals

Purity of SAMPL6 pKa challenge compounds were determined based on LC-MS. The purity analysis was performed using an Agilent HPLC1200 Series equipped with auto-sampler, UV diode array detector and a Quadrupole MS detector 6140. The software used was Chemstation for LC & LCMS with version C01.07SR2. The column for the analysis is Assentis Express C18 3.0x100mm 2.7 μ l particle size at Column temperature = 45°C.

- Mobile phase A: 2 mM ammonium formate (pH=3.5) aqueous
- Mobile phase B: 2 mM ammonium formate in Acetonitrile : Water=90:10 (pH=3.5)
- Flow rate : 0.75ml/min
- Gradient: Starting with 10% B to 95%B in 10 minutes then hold 95%B for 5 minutes.
- Post run length: 5 minutes
- Mass condition: ESI positive and negative mode
- Capillary voltage: 3000 V
- Drying gas flow: 12 ml/min
- Nebulizer pressure: 35psi
- Drying temperature: 350°C
- Mass range: 5-1350 Da; Fraggmentor:70; Threshold:100

The percent area for main peak is calculated based on the area of main peak divided by the total area of all peaks. The percent area of the main peak is reported as an estimate of sample purity.

SI Figure: all MS data

Results

TABLE: Measured macroscopic pKa for all compounds, what method

Effect of cosolvent on measured pKas

FIGURE: Cosolvent vs non-cosolvent measurement difference

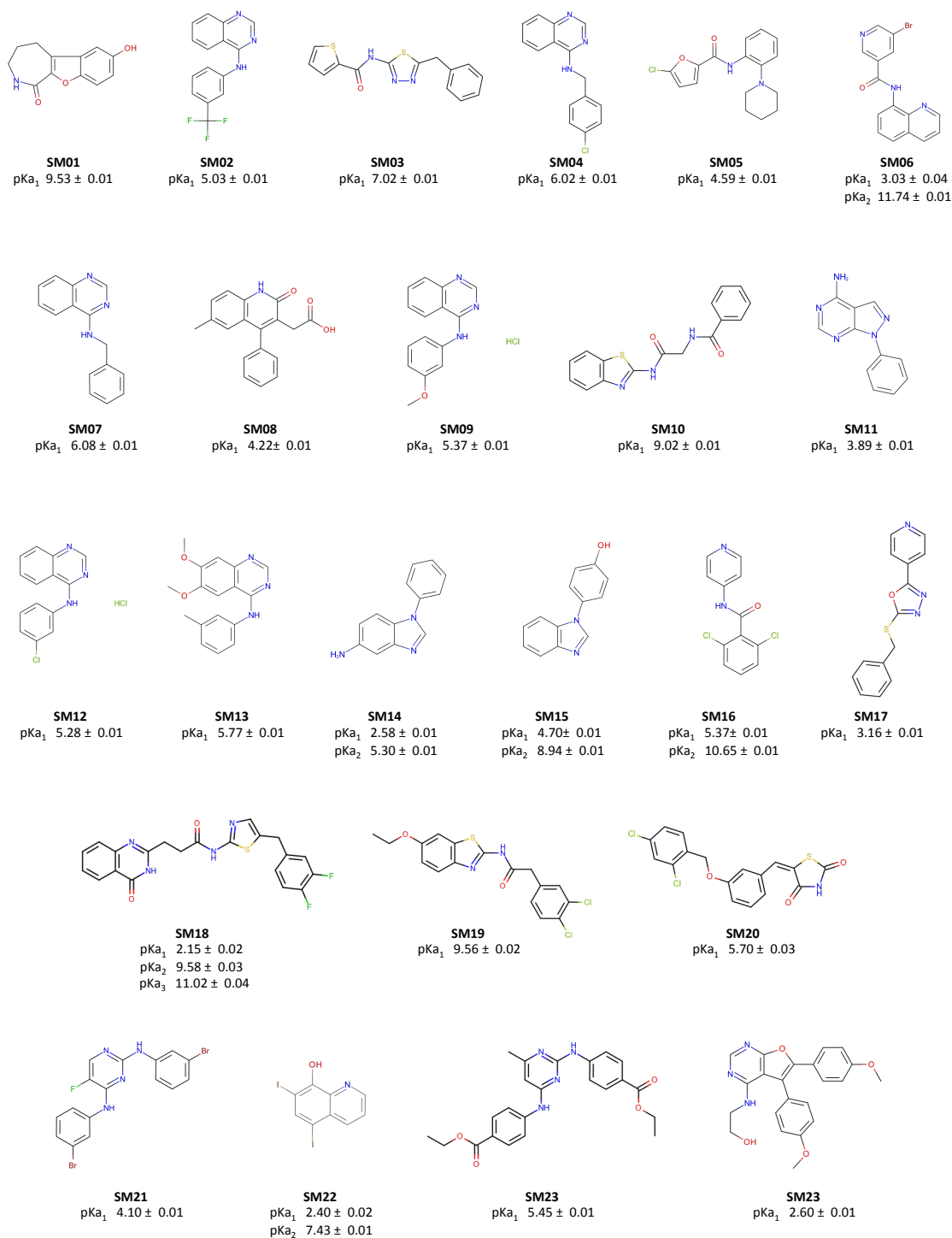


Figure 1. 24 molecules of SAMPL6 pKa challenge and their experimental pKa values measured with UV-metric method of Sirius T3. Uncertainties of pKa measurements are expressed as SEM calculated from three independent measurements.

Discussion

Discussion of experimental data interpretation: macroscopic Multiwavelength absorbance analysis on the Sirius T3 allows for very good resolution of pKas, but it is important to note that this method produces estimates of macroscopic pKas. If multiple microscopic pKas have close pKa values and overlapping changes in UV absorbance spectra associated with protonation/deprotonation event, the spectral analysis could produce a single macroscopic pKa that represents an aggregation of multiple microscopic pKas. Lessons learned for future challenge iterations

Challenges with NMR measurements?

pKa can shift based on ionic strength, temperature, and lipophilic content

Can other cosolvents be interesting for process chemistry.

Code and data availability

- SAMPL6 pKa challenge instructions, submissions, and analysis is available at <https://github.com/MobleyLab/SAMPL6>
- Python scripts used for compound selection are available at **compound_selection** directory of <https://github.com/choderalab/sampl6-physicochemical-properties>

Construct a proper README for compound selection directory.

Overview of supplementary information

Author Contributions

Complete this section.

Conceptualization, ; Methodology, ; Software, ; Formal Analysis, ; Investigation, MI, ; Resources, ; Data Curation, ; Writing-Original Draft, ; Writing - Review and Editing, ; Visualization, ; Supervision, ; Project Administration, ; Funding Acquisition,

Acknowledgments

Complete this section.

MI, ASR, and JDC acknowledge support from the Sloan Kettering Institute. JDC acknowledges support from NIH grant P30 CA008748. MI acknowledges Doris J. Hutchinson Fellowship. Acknowledging Merck Preformulation department for materials, expertise and instrument time.

Brad Sherborne Paul Czodrowski - helped with purchasable compound list in earlier iteration Ikenna Ndukwe Caitlin Bannan

Disclosures

JDC is a member of the Scientific Advisory Board for Schrödinger, LLC.

References

- Tam KY, Takács-Novák K. Multi-wavelength spectrophotometric determination of acid dissociation constants: a validation study. *Analytica chimica acta*. 2001; 434(1):157–167.
- Allen RI, Box KJ, Comer JEA, Peake C, Tam KY. Multiwavelength spectrophotometric determination of acid dissociation constants of ionizable drugs. *Journal of pharmaceutical and biomedical analysis*. 1998; 17(4):699–712.
- Manallack DT. The pK(a) Distribution of Drugs: Application to Drug Discovery. *Perspectives in Medicinal Chemistry*. 2007 Sep; 1:25–38.
- Avdeef A, Box KJ, Comer JEA, Gilges M, Hadley M, Hibbert C, Patterson W, Tam KY. PH-metric logP 11. pKa determination of water-insoluble drugs in organic solvent-water mixtures. *Journal of pharmaceutical and biomedical analysis*. 1999; 20(4):631–641.

159 **Supplementary Information**